Summary of the ASCI/NNSA Verification Workshop 2001

Overview

A Verification Workshop, sponsored by the NNSA and the LANL V&V program, was held at LANL on 28–29 November 2001. The impetus for the workshop was provided by the ASCI V&V executives: Richard Klein (LLNL), Ken Koch (LANL), Cynthia Nitta (LLNL), and Marty Pilch (SNL).

The Verification Workshop, which consisted of over a dozen technical presentations—many related to the Tri-Lab Performance and Engineering Verification Test Suites—and roundtable discussions, was attended by over 40 representatives of the engineering and performance communities at LANL, LLNL, SNL, and NNSA. In addition, there were three keynote talks: the first by Tim Trucano (SNL), who provided an overview of V&V; the second by Matt Kirkland (LANL), who discussed the LANL primary validation strategy; and the third by Chris Morris (LANL), who provided the audience with an introduction to proton radiography.

This summary document consists of the following three sections.

1. A synopsis that documents the key topics regarding verification discussed at this workshop. This synopsis does not review the substance of the technical presentations; rather, it provides a record of some of the issues raised during the course of the meeting. The issues presented here were distilled from:
   (i) the opening remarks of Bill Rider;
   (ii) the roundtable discussions moderated by Jim Kamm, Richard Klein, Roger Logan, and Bill Rider; and
   (iii) the summary observations of Tim Trucano.
2. A copy of the workshop agenda.
3. A list of the registered workshop attendees.

We provide this summary document as information to the larger NW community. If you would like further information about either this Workshop or the verification activities discussed herein, we encourage you to contact the Workshop organizers, Jim Kamm (kammj@lanl.gov, 505-667-1918) and Bill Rider (wjr@lanl.gov, 505-665-4162), or the responsible parties at your host institution.

I. Verification Workshop Technical Synopsis

A. The "Seven Deadly Sins and Seven Virtuous Practices" of V&V

Bill Rider (LANL), Jim Kamm (LANL)

Verification and validation help ensure quality by demonstrating codes' strengths and weaknesses. These practices both establish code capabilities and assist in the allocation of resources for additional code development. To be a meaningful endeavor, V&V should be rigorous, systematic, and self-consistent; occasionally, however, what passes for V&V does not meet these criteria. We acknowledge that we ourselves—as code developers and V&V analysts—are guilty of certain "deadly sins" listed below. The "sinful" nature of some of these activities, however, is manifested when one is content with only these practices. Each of us is aware of the resources, effort, and (frankly) drudgery involved in performing systematic and quantitative V&V, the results of which are sometimes imperfect. We contend that the "virtuous practices" listed below encompass many of the necessary (but not sufficient) characteristics of V&V that are required to provide less ambiguous statements of codes' capabilities. As a community, we should strive to enhance the "virtuous" nature of our V&V activities.

	"Deadly Sin"	*"Virtuous Practice"*
1.	Assume the code is correct	Assume the code has flaws—and find them!
2.	Qualitative comparison	Quantitative comparison
3.	Use of problem-specific settings	Verify & validate the same settings and code
4.	Code-to-code comparisons only	Use analytic solutions and experimental data
5.	Computing on one mesh only	Systematic mesh refinement
6.	Show only results that make the code "look good"	Show all results to highlight strengths and reveal shortcomings
7.	Don't differentiate between accuracy and robustness	Assess accuracy and robustness separately

B. Roundtable Topics: Future Verification Activities

Jim Kamm (LANL), Richard Klein (LLNL), Roger Logan (LLNL), Bill Rider (LANL)

After the technical presentations, there were three roundtable discussions, moderated by the above individuals. The purpose of these sessions was to discuss issues touched on, but not fully explored or resolved, during the technical presentations. In the list below, we have attempted to summarize the most significant points raised in those discussions. This list may provide guideposts for the improvement of ongoing and future verification activities.

- Demonstration of code convergence on smooth problems
  - *Establishing code performance on smooth problems ensures that the underlying algorithms are performing as designed in ideal (i.e., the least demanding) circumstances.*
- Analysis of each problem at more than a single simulation time
  - *Analysis of multiple "snapshots" may reveal additional (unexpected?) aspects of code behavior and, therefore, is desirable.*
- Perform both spatial and temporal convergence analysis
  - *Both space-evolution and time-advancement algorithms should be independently verified.*
- Investigate the implications of degraded convergence of high-order schemes
  - *One should (i) demonstrate that high-order schemes are "worth it" and (ii) assess the effects of high-order schemes on strongly nonlinear problems.*
- More rigorous justification of the "verification suite", with the possible addition of supplementary or better delineated verification problems
  - *The engineering community demonstrated a greater number of regression and verification problems than the performance community typically considers.*
  - *Should it be incumbent upon the individual who specifies a verification problem to also specify the acceptance criteria for its successful calculation?*
  - *What additional (i.e., more "realistic", more "useful") problems can we incorporate into future test suites?*
- Extension of verification activities to ALE and AMR technologies
  - *How should mesh refinement studies of these technologies be conducted?*
  - *How should these technologies themselves be verified?*
- Facilitate the interaction between verification to validation
  - *What is the proper emphasis in the continuum from "code physics verification" to "calculation verification" to "code physics validation" to "code-to-code comparison"?*
  - *What are the appropriate analyses to be conducted in each of these endeavors?*
- Actively promote interaction with the design community
  - *A more diverse set of verification problems (thank you, Bill Chandler!) together with increased person-to-person interaction will help build alliances and break down barriers.*
- Learn from different communities' V&V standards and practices
  - *E.g., software vendors' extensive regression, verification activities, greater use of standards, and providing code support engineers: what can or should we incorporate in our practices?*
  - *E.g., examining and learning from the IEEE standards may assist our establishment of improved V&V practices.*
  - *E.g., Roger Logan and Cynthia Nitta's "V+V meters" and "ladder diagrams" give clear, concise information on the maturity and evolution of code capabilities.*
- Promote validation activities in a comparable Validation Workshop
  - *Code developers and verification analysts must contribute to and actively participate in Validation Workshop discussions!*

C. Workshop Summary: Issues for Thought, Discussion, and Further Consideration
Tim Trucano (SNL)

The concluding talk at the workshop was an overview of the workshop itself, kindly provided by Tim Trucano (SNL), who had offered his own knowledgeable perspectives on V&V in the kickoff keynote presentation. Below are some of the strategic observations that Tim shared with the attendees regarding the content and direction of the workshop.

- How to develop consensus regarding code "qualification", i.e., the criteria that must be satisfied so that the community agrees the code is "good enough"
  – *We must not lose sight of the guiding notion behind verification: that we get the right answers for the right reasons—and are able to demonstrate it!*
  – *To achieve this goal, the community must develop responses to the fundamental questions "What defines success?" and "What defines failure?" for the verification suite.*
  – *There are no "magic bullets" in this process—and simply accelerating the times at which V&V victory is to be declared may be trying to achieve a "magic bullet"!*

- Verification Suites - and what the community should define about them:
  (i) Why are these test problems/this test suite needed?
  (ii) How is the test suite populated with tests?
  (iii) What is the result of the tests/what do these results mean?

  The workshop provided significant discussion on why test problems were selected and how they were populated (i.e., defined); however, little direct discussion on assessment was presented. By this, Tim meant that while comparison metrics were discussed, explicit determination of "Pass/Fail" or "Success/Failure" criteria was typically missing. This is not surprising because this task is not trivial. Nonetheless, assessment criteria that include decisions about "Pass/Fail" together with the definition of comparison metrics are critical for several reasons, including the following.

  - They contribute to an environment of verification "standards" that have broad relevance, understanding, and consequence in the M&S community.
  - Accurate and quantitative assessment criteria are required (REQUIRED!) to precisely extrapolate the confidence we achieve from the narrow confines of verification test suites to the broader goals of validation and application.
  - Accurate and quantitative assessment criteria are required (REQUIRED!) for qualification.

  The "Pass/Fail" aspects of assessment were indirectly discussed at the workshop, as reflected by the discussion of the relevance of several of the test problems as currently defined. This discussion was typically along the lines of "For application *A* do you really care that you calculate feature *F* of test problem *T* to within the stated accuracy?" This is precisely an example of what Tim is thinking about in terms of "Pass/Fail" criteria. By one criterion the simulation of the test problem might achieve a "Pass", but by another it might achieve a "Fail." The metric did not change—the goal of the problem did! It is essential that these criteria be coupled to the "Why" of the test problem suite; these criteria likely influence the "How", as well. Addressing the meaning of "Pass/Fail" only AFTER calculations have been performed is not likely to accomplish the goals of verification.

  Tim also commented that there are other technical methodologies available for addressing the "How" factor that were not discussed at all at the Workshop. These include statistical testing based on approximations to the expected code paths that would be executed during "real" applications. Tim feels that statistical testing methodologies, as well as broader understanding

of probabilistic software reliability, are fruitful avenues for future verification work. Firmer assessment criteria—in particular, how to define "failures"—are required for this, however.

- Code-to-code comparisons (C2CC)
  – *When (i.e., at what stage of code development) should C2CC be done?*
  – *How (e.g., with what metrics, on what problems) should C2CC be done?*
  – *Can C2CC indeed even be done in a meaningful fashion?*
  – *What is the utility of having "gold standard" solutions to certain problems?*
  – *Are such numerically computed "gold standard" solutions even achievable?*

- How can we achieve integrated testing that conveys relevance to the designers?

  There was concern expressed about the apparent widening gap between the kind of painstaking verification work that was presented at the Workshop (and which certainly needs to be performed) and the NW community's need for short timescale validation, which is followed by the even shorter timescale of the NW application. Tim pointed out that the best way to define and implement a V&V program is to have carefully crafted and COUPLED activities simultaneously in verification, validation, and exploratory applications. This is recognized by the software engineering community and is a major reason that the standard estimates for the cost of VV&A (using the DoD DMSO acronym) are two to three TIMES the cost of writing the software. Needless to say, this kind of money is not (currently) available for ASCI-related V&V. The fact remains that Tim emphasized the COUPLING of V&V. Choosing to do only verification is as bankrupt a choice as choosing to do only validation, and both—together—make more sense when performed in conjunction with carefully chosen and limited attempts at applications. How to achieve this balance with the resources available to the ASCI V&V program probably remains one of the greatest challenges that we face.

- How can verification activities better interface with code projects?

  Tim stressed the following point regarding "Independent V&V." If by "Independent" we mean "Hostile and Critical", then we are *doomed to failure*. Tim prefers to interpret "Independent" as "Integrated But Separate." The meaning of this important choice of terminology is described below.

  - V&V should be integrated with the code project as early as possible, preferably beginning at the initial design phases of the code.
  - Specific V&V analysts IN ADDITION TO code developers should be defined and woven into the fabric of the project. A typical characterization of these roles is "beta testing" (Tim jokingly referred to them as "victim users" at the Workshop.) The key emphasis, though, is that these roles are viewed as an INTEGRAL PART OF the code project, neither disjoint from nor hostile to the project.
  - This integration requires planning and coordination with the software development function(s). For example, it makes little sense to "prove" through verification that a capability that has not been implemented does not work. On the other hand, design of verification activities for a software capability should not have to wait until the capability has been deployed. All three of the key elements in verification activities—(i) goal of the testing, (ii) design of the testing, (iii) assessment of the testing—can be defined prior to implementation. Indeed, the earlier and closer the interaction between V&V and the code development effort, the more likely these key elements are to be better defined.
  - Obviously, the reward system influences the success of this integration. It is not a desirable end product to have code developers who are criticized (i.e., "slaughtered") because verification reveals bugs. On the other hand, it is also not desirable to have "beta testers" marginalized (if not outright ostracized) because their activities reveal weaknesses in im-

plemented capabilities. Taking to heart the "Seven Deadly Sins and Seven Virtuous Practices" that Bill Rider emphasized can help eliminate these undesirable consequences.

Tim doubts the efficacy of the following approach to V&V: "A period of time *T* passes, then an 'independent V&V' effort begins." Rather, Tim envisions a staged process that better reflects integrated yet separate roles and responsibilities:

- "V&V" is involved from the BEGINNING of the code project;
- Specialists (a.k.a. beta testers or "victim users") are involved from the beginning and integrated into the code development project;
- The approach is then: "After a period of time *T*, during which phased and increasing V&V has been performed, start growing a larger user community, with the understanding that the initial emphasis of this community is STILL in V&V and carefully designed and constrained applications. Aim to deliver the code capability to the NW application community for higher importance applications after time *T+dT*." This approach is not unlike traditional code development at the labs. Tim's opinion, however, is that the greater formality implied lends itself to time scales that are overall shorter than the historical development and release cycles for complex scientific software.

Tim's point about "test engineers" was intended to be more of a metaphor for having responsibility built into the code team to handle the growth in V&V testing and its consequences. This growth will inevitably occur if V&V is well integrated into the project. This person(s) is a logical interface between individuals whose roles are designing algorithms and writing code and people whose primary role is to perform separate V&V.

- Where are we going as a community?

Tim pointed out that these workshops offer an important vehicle for collaboration with the NW community. Tim believes that the assumption that a "V&V-ed Code" (or even just a "Verified Code") will suddenly be delivered at some fixed date for NW applications is an extraordinarily dangerous and naïve act-of-faith. Continued workshops with focus on verification, validation, and technical requirements of NW work that include the ENTIRE community (code developers, V&V practitioners, and NW designers) will dispel this notion and elaborate the more realistic and helpful notion of EVOLUTION of QUALIFIED code capabilities. We must not lose sight of the bottom line: *to make a positive impact on the stockpile through the ASCI codes*.

II. Verification Workshop Agenda

Day 1 Wednesday 28 November 2001
The Agnew Room (TA-3, SM 43, Rm A164)

8:30 No-host Continental Breakfast
Juice, coffee, fruit, pastries, breakfast burritos

9:00 Welcome + Operational Remarks J.Kamm/W.Rider
Meeting participants will be encouraged to engage in frank exchanges on the technical content, research directions, and programmatic impact of V&V.

9:05 Opening Remarks D.Shirk
Why V&V, in general, and verification, in particular, are important to DOE DP activities.

9:15 Program Overview M.Pilch
How does the V&V program dovetail with other DP activities?

9:30 Keynote 1: Overview of Verification & Validation T.Trucano
This talk is intended to establish a common language on V&V for the workshop attendees.

10:15 Break

The next two talks introduce the problem sets, analysis techniques, as well as their relevance to the programmatic applications.

10:30 Introduction/Overview of Performance Side Verification R.Klein

11:00 Introduction/Overview of Engineering Side Verification D.Crane

11:30 Lunch (Otowi/Motorola Research Park)

The foundation laid in the next two sessions will provide the common language and template for the workshop discussions that will follow. The point of these talks is not only to discuss the verification problems, but also to present the <u>details</u> of the analysis method, the <u>details</u> of the results, and the <u>implications</u> of those results.

13:00 Sedov & Noh Problem Descriptions, Code Results, Analysis, Issues
3/4 hour LLNL [B.Moran], 3/4 hour LANL [J.Kamm, M.Alme]

14:30 "Prototypical" Engineering Problem Descriptions, Code Results, Analysis, Issues
LANL, LLNL, SNLA D.Sam

15:30 Break

16:00 Keynote 2 Verification & Validation in Related Areas
LANL Validation Process [M.Kirkland] & Proton Radiography [C.Morris]
The point of these presentations is to illustrate selected programmatic directions that can be meaningfully impacted by V&V activities.

17:00 Operational Remarks + Overview of Next Day J.Kamm/W.Rider

18:30 Informal Social Dinner — Gabriel's (in Cuyamungue), tel. 455-7000
The intention is to bring workshop participants together in an off-site social setting.

Day 2 Thursday 29 November 2001
The Agnew Room (TA-3, SM 43, Rm A164)

8:00 No-host Continental Breakfast
Juice, coffee, fruit, pastries, breakfast burritos

8:25 Operational Remarks + Overview of the day J.Kamm/W.Rider
Workshop participants will again be encouraged to actively participate in the day's discussions.

8:30 Additional performance verification problems
1/2 hour LLNL [S.Brandon],
1/2 hour LANL [S.White/C.Royer, J.Painter, M.Clover],
Discussion of verification problems, including the Tri-Lab test suite together with other problems, that expose unresolved issues in verification or code capabilities; specifically, we hope to discuss problems (1) that are difficult to set-up, (2) for which the "exact" solution is difficult to evaluate, and (3) for which the comparison of the code-solution and "exact" solution is problematic.

9:30 Engineering Code Projects Responses to Verification Results
1/2 hour SNLA [J.Mitchell], 1/2 hour LLNL [A.Anderson]
These presentations will discuss the status of engineering verification test suite problems, how code teams have responded to these results, and the effect that these results have had on the code development process.

10:30 Break

10:45 Performance Code Projects Responses to Verification Results
1/2 hour LLNL [S.Brandon], 1/2 hour LANL [M.Alme, J.Painter, M.Clover]
These presentations will discuss the status of performance verification test suite problems, how code teams have responded to these results, and the effect that these results have had on the code development process.

11:45 Lunch (Otowi/Motorola Research Park)

13:00 Roundtable 1: Tie Up Loose Ends from Verification Problems
Moderators: J.Kamm & R.Klein
This discussion will address outstanding "big picture" issues raised earlier in the workshop.

The following two discussions address the following two questions:
1. *Where does verification go from here?*
2. *How can we tie verification to programmatic objectives in a more meaningful way?*

14:00 Roundtable 2: Performance V&V discussion: future directions and relevance to programmatic activities
Moderator: W.Rider

14:45 Roundtable 3: Engineering V&V discussion: future directions and relevance to programmatic activities
Moderator: R.Logan

15:30 Concluding Remarks T. Trucano
These remarks will summarize what has been accomplished in the workshop, and put forth suggestions on how to proceed in a productive manner.

15:45 Adjourn

III. Verification Workshop Registered Attendees

Affiliation	*Name*	*Email*	*Telephone*
LANL	Marv Alme	ALME@LANL.GOV	505-665-3869
	Jerry Brock	JSBROCK@LANL.GOV	505-665-3210
	Dominic Cagliostro	DJC@LANL.GOV	505-667-8500
	John Cerutti	JHC@LANL.GOV	505-667-0378
	Bill Chandler	CHANDLER@LANL.GOV	505-665-9123
	Larry Cox	LJCOX@LANL.GOV	505-665-7344
	David Crane	DCRANE@LANL.GOV	505-665-7245
	Barbara Devolder	BGD@LANL.GOV	505-667-8928
	Ron Dolin	RMD@LANL.GOV	505-667-9142
	Mabel P Grey-Vigil	MPG@LANL.GOV	505-665-1318
	Alexandra Heath	ARH@LANL.GOV	505-667-2558
	Jim Kamm	KAMMJ@LANL.GOV	505-667-1918
	Joe Kindel	JKINDEL@LANL.GOV	505-667-7299
	Matt Kirkland	MKIRKLAND@LANL.GOV	505-667-3105
	Ken Koch	KRK@LANL.GOV	505-667-4288
	Douglas B Kothe	DBK@LANL.GOV	505-667-9089
	Chuck Lebeda	CFL@LANL.GOV	505-667-8120
	Mary Beth Lee	MBL@LANL.GOV	505-665-4099
	Christopher Morris	CMORRIS@LANL.GOV	505-667-5652
	Jim Painter	JWP@LANL.GOV	505-667-7658
	Doug Post	POST@LANL.GOV	505-665-7680
	Bill Rider	WJR@LANL.GOV	505-665-4162
	Cheryl L Royer	CROYER@LANL.GOV	505-665-2154
	Don Shirk	DGS@LANL.GOV	505-667-0940
	David Tubbs	DLT@LANL.GOV	505-667-7745
	Bob Webster	ROBW@LANL.GOV	505-665-4183
	Dan Weeks	WEEKS@LANL.GOV	505-665-8691
	Steve White	SWHITE@LANL.GOV	505-667-4623
LLNL	Andy Anderson	ANDERSON1@LLNL.GOV	925-423-9634
	John Bolstad	BOLSTAD1@LLNL.GOV	925-423-2485
	Scott Brandon	BRANDON1@LLNL.GOV	925-423-6804
	Richard Klein	KLEIN4@LLNL.GOV	925-422-3548
	Dave Kraybill	KRAYBILL1@LLNL.GOV	925-422-4674
	Roger Logan	RWLOGAN@LLNL.GOV	925-423-4872
	Bill Moran	MORAN1@LLNL.GOV	925-422-7250
	Cynthia Nitta	NITTA1@LLNL.GOV	925-423-3792
	Doug Peters	PETERS3@LLNL.GOV	925-424-3554
	David Sam	SAM1@LLNL.GOV	925-422-4149
	Tod Woods	TOD@EAGLEVAIL.LLNL.GOV	925-422-6957
SNLA	Steve Lott	SELOTT@SANDIA.GOV	505-844-1889
	John Mitchell	JAMITCH@SANDIA.GOV	505-844-3435
	Martin Pilch	MPILCH@SANDIA.GOV	505-845-3047
	John Pott	JPOTT@SANDIA.GOV	505-284-3591
	Chris Roy	CJROY@SANDIA.GOV	505-844-9904
	Timothy G. Trucano	TGTRUCA@SANDIA.GOV	505-844-8812
DOE	Aamer Haque	AAMER.HAQUE@NNSA.DOE.GOV	202-586-7289
	Ed Pate	BRUCE.PATE@NNSA.DOE.GOV	202-586-5630
	Jamileh Soudah	JAMILAH.SOUDAH@NNSA.DOE.GOV	202-586-2104
SAIC	Mike Clover	MICHAEL.R.CLOVER@SAIC.COM	858-826-5926